Ground-state properties of crystalline silicon in a density-functional pseudopotential approach

Alex Zunger
Solar Energy Research Institute, Golden, Colorado 80401
(Received 7 August 1979)

A first-principles nonlocal pseudopotential approach is shown for the first time to predict accurately the ground-state bulk properties of a semiconductor. The calculated equilibrium lattice constant, total valence energy, and bulk modulus of Si are within 0.2%, 0.5%, and 5%, respectively, of the observed values.

I. INTRODUCTION

While many of the ground-state electronic properties of simple and transition-metal elemental solids are quantitatively well understood within the local-density formalism (DF), both in its all-electron (AE) and valence-only (pseudopotential) versions, similar nonempirical studies of the opened structure, covalently bonded solids are scarce. For semiconductors in particular, no self-consistent nonempirical calculation of the ground-state properties exists. For Si, empirical (local) pseudopotentials fitted to the spectra of either solids or bare ions and used successfully to predict the one-electron band structure of semiconductor surfaces, chemisorption, and defects have so far yielded small equilibrium bond lengths for the bulk solid and for the Si molecule. Furthermore, these pseudopotentials have produced systematic errors in the topology of the ground-state charge density (i.e., the ellipsoidal density contours in the bulk solid are perpendicular, rather than parallel to the bond axis) as well as significant deviations (70%) between the calculated and measured bulk modulus. Empirical tight-binding methods can be parametrized to successfully predict geometries at semiconductor surfaces but fail in predicting structural information even for the simplest bulk solids. In this paper I show for the first time that a self-consistent approach using a first-principles density-functional (nonlocal) pseudopotential predicts remarkably well the equilibrium lattice constant, total valence energy, bulk modulus, and the ground-state charge density of bulk Si (i.e., with errors of 0.2%, 0.5%, and 5% for \(a_{\text{eq}}\), \(E_{\text{tot}}\), and \(B\), respectively). The origin of the discrepancies previously obtained with the empirical pseudopotentials hence does not lie in the pseudopotential formalism or in the local-density formalism (used to screen these potentials), but rather in the approach adopted to construct the pseudopotentials.

In contrast with previous model pseudopotentials, the present pseudopotential is derived from a well-defined electronic-interaction model (the density-functional formalism) rather than from an empirical adjustment. The implications of this are: (i) one can clearly assess the effects of a chosen representation of exchange and correlation screening potential on the ground-state properties and (ii) a contact is made with the all-electron (i.e., treating core plus valence wave functions on the same footing) theories which are based on the same electronic interaction model: at the limit where the underlying pseudopotential frozen-core approximation is exact, the pseudopotential results should become identical with the AE results. In contrast with the AE approach however, the present method has the advantage of projecting out from the total energy the large (98%) and structurally insensitive part associated with the nearly inert core electrons. Hence, while the cohesive energy of crystalline Si, Ge, and Sn as a fraction of the total AE energy per atom is as small as \(5.5 \times 10^{-4}\), \(6.2 \times 10^{-5}\), and \(1.7 \times 10^{-5}\), respectively, the same cohesive energy expressed as a fraction of the total pseudopotential energy per atom is \(4 \times 10^{-2}\), \(3.4 \times 10^{-2}\), and \(3.2 \times 10^{-2}\), respectively. This increase by two to three orders of magnitudes in the relative energies, coupled with the high precision demonstrated here in calculating the pseudopotential total energy makes it hopeful that the present approach may come close to predicting quantitatively structural energies of complex systems as well as revealing systematic trends in physical and chemical ground-state properties of classes of solids.

II. METHODOLOGY AND RESULTS

The first-principles pseudopotentials are derived from a density-functional (DF) formulation of a pseudoatom interacting with an external potential \(V_{\text{f}}(r)\). A variational procedure which includes a number of Lagrange-multiplier constraints designed to produce a relation between the spectral properties and charge density of the pseudoatom and the real atom is used to solve for the nonlocal pseudopotential \(V_{\text{f}}(r)\). This potential differs from the empirical potentials \(V_{\text{emp}}(r)\) used extensively in the literature.
(e.g., Ref. 7 and references therein) in that (cf. Fig. 1) \( V_l(r) \) (i) is angularly nonlocal, reflecting the restrictions imposed by the Pauli principle, (ii) it assumes its asymptotic value of \(-4/r\) faster than \( V_{\text{emp}}(r) \), and (iii) it is significantly "stronger" than the empirical potentials. These high-momentum \((q)\) transfer components of \( V_l(q) \) which are absent in \( V_{\text{emp}}(q) \) are a direct consequence of the imposition of a minimum pseudowave-function core-amplitude condition and the "maximum-similarity" constraint between the "true" and the pseudo wave functions, in the chemically important valence region.\(^{14}\) The latter condition assures that the chemical information included in the tail part of the "true" atomic valence wave functions is contained in the pseudo wave functions. The former conditions (cf. minimization of the pseudopotential energy dependence) acts to increase the likelihood that this continues to be a good approximation even if \( V_l(r) \) replaces the effect of the same core in systems other than the atom (e.g., molecule, solid, etc.).

A central point in deriving our pseudopotential is that we first construct a pseudo wave function which has certain imposed relations to the "true" all-electron valence wave function and then solve for the pseudopotential which when screened self-consistently, will produce these wave functions as ground-state eigenstates with energy eigenvalues which equal those obtained in the full all-electron calculation. In contrast, the empirical as well as semiempirical pseudopotential approaches assume a parametric ansatz form for the pseudopotential and vary its parameters to obtain agreement with certain experimental data, leaving the pseudo wave functions to be largely determined implicitly by such an energy-fitting procedure.

Our pseudo wave functions \( \{ \psi_{\text{ps}}(r) \} \) are represented as a sum of core and valence orbitals of the all-electron density-function problem \( \{ \psi_{\text{el}}(r) \} \) plus a wave-function component \( f_{\text{ps}}(r) \) which lies outside the core plus valence DF orbital space.\(^{14}(b)\) For Si we obtain

\[
\begin{align*}
X_{3s}(r) &= C_{3s}\Psi_{3s}(r) + C_{2s}\Psi_{2s}(r) + \psi_{3s}(r) + A_{3s}r^2e^{-\alpha r}, \\
X_{3p}(r) &= C_{3p}\Psi_{3p}(r) + \psi_{3p}(r) + A_{3p}r^2e^{-\alpha r'}, \\
X_{3d}(r) &= \psi_{3d}(r).
\end{align*}
\]

The coefficients \( \{ C_{Al, \beta, \alpha_i} \} \) are determined such that: (i) the pseudo wave functions are normalized, monotonic, and nodeless; (ii) the pseudo wave functions have a minimum core amplitude with \( X_{\text{ps}}(0) = X_{\text{el}}'(0) = \chi_{\text{el}}'(0) = 0;\) (iii) there is a maximum similarity between the "true" all-electron valence wave functions and the pseudo wave functions \( (\text{i.e., } X_{3s} \text{ and } \psi_{3s}, \text{ } X_{3p} \text{ and } \psi_{3p}, \text{ and } X_{3d} \text{ and } \psi_{3d}) \) starting from \( r = \infty \) and going inwards to the smallest \( r = R_e \) value possible under constraints (i) and (ii).

As there are no core states of \( I = 2 \) symmetry in Si, the all-electron wave function \( \psi_{3d}(r) \) satisfies condition (i) and hence \( X_{3d}(r) = \psi_{3d}(r) \).

The coefficients \( \{ C_{Al, \beta, \alpha_i} \} \) are given in Table I where the exchange parameter \( \alpha = \frac{1}{3} \) and the electronic configuration \( 1s^22s^22p^63s^13p^23d^0 \) was used in solving the all-electron problem for \( \{ \psi_{\text{el}} \} \). Our pseudo wave functions are identical to the "true" all-electron wave functions asymptotically at large \( r \).

**TABLE I.** Coefficients of the pseudo wave functions in Eq. (1), in atomic units. The all-electron orbitals \( \psi_{\text{el}}(r) \) are obtained from a density-functional solution with the configuration \( 1s^22s^22p^63s^13p^23d^0 \) and exchange coefficient \( \alpha = \frac{1}{3} \). The orbital energies of the pseudo- as well as all-electron eigenvalue problem are: \( \epsilon_{3s} = -0.48805 \text{ a.u.}, \epsilon_{3p} = -0.22987 \text{ a.u.}, \) and \( \epsilon_{3d} = -0.007324 \text{ a.u.} \) The crossing points \( V_{\text{ps}}(R_e) = 0 \) of the bare pseudopotential, including the centrifugal term \( l(l+1)/2r^2 \) are: \( R_e = 0.6847, \ ) R_\beta = 0.6679, and \( R_\delta = 0.5342 \text{ a.u.} \) The total energy of the pseudoatom, solved self-consistently in the \( 3s^13p^23d^0 \) configuration is \(-3.244726 \text{ a.u.}\) Not that \( \psi_{3s}(r) \) is positive while \( \psi_{3p}(r) \) is negative.

<table>
<thead>
<tr>
<th>3s orbital</th>
<th>3p orbital</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_{1s} ) = 0.011202 772</td>
<td>( C_{2p} ) = -0.252 506 008</td>
</tr>
<tr>
<td>( C_{2s} ) = -0.339 808 307</td>
<td>( A_{3p} ) = 4.257 435</td>
</tr>
<tr>
<td>( A_{3s} ) = -4.711 926</td>
<td>( \alpha_3 ) = 3.85</td>
</tr>
</tbody>
</table>

*Fig. 1.* Empirical local potential \( V_{\text{emp}} \) [Ref. 6(a)] and the first-principles nonlocal potentials \( V_l \). Also shown are the all-electron 3s orbital \( X_{Al}(r) \) (---) and the pseudo-orbitals obtained with the empirical (---) and first-principles (- - -) pseudopotentials \( \{ X_{\text{emp}}(r) \text{ and } X_{\text{fp}}(r) \}, \text{ respectively} \).
have a charge accumulation function

\[ Q_{\text{cal}}(R) = \int_0^R |x_n|^2 \, dr \]

which equals the all-electron one

\[ Q_{\text{cal}}^A(R) = \int_0^R |\psi_n|^2 \, dr \]

from \( R = \infty \) to a point inward of the last maxima of \( \psi_n(r) \) (and hence both sets of wave functions produce an identical electrostatic field in this region). In addition, the low core amplitude of \( x_n(r) \) ensures a low-energy dependence (and hence high transferability) of the resulting nonlocal pseudopotentials.\(^{14}\)

As seen in Fig. 1, the empirical pseudopotential \( V_{\text{emp}} \) produces\(^{6(a)}\)\(^{7}\) in the atom a significant charge density in the core region at the expense of reducing it in the valence region. When applied to crystalline Si, or to the Si\(_2\) molecule, \( V_{\text{emp}} \) leads to a bond charge which is substantially different from either the experimental result or the full all-electron calculation.\(^{12}\) We find that this reduction in the valence charge density leads to a lowering of the interelectronic Coulomb repulsion energy. Coupled with the over attractiveness of the non-Coulombic part of \( V_{\text{emp}} \) at the bond center (cf. Fig. 1), this leads to an anomalously small equilibrium lattice constant [e.g., 4.7 (Ref. 9)–5.3 (Ref. 10) Å instead of the observed value 5.43 Å]. In the case of the empirical potential of Ref. 6(a), this effect also yields an overly attractive total valence energy [e.g., about \(-8.42\) Ry/atom (Ref. 9) relative to the observed value of \(-7.9\) Ry/atom]. The empirical potential of Ref. 6(b) being somewhat less attractive at the bond center, was shown\(^{10}\) to yield the correct valence energy. Similar conclusions are born out by the comparison of the empirical pseudopotential and the all-electron density-functional results for Si\(_2\).\(^{11}\)

In the present approach we use the density-functional pseudopotential \( V_f(r) \) to solve self-consistently for the Si band structure as a function of the lattice parameter and to compute the total ground-state energy. In a recent calculation,\(^{16}\) the present Si first-principles pseudopotential was shown to reproduce the all-electron DF self-consistent bulk band structure to within an average deviation of 0.06 eV/state over an energy range of 20 eV.

The effective potential we use is a screened nonlocal pseudopotential

\[
V_{\text{eff}}(\rho(\vec{r})) = \sum_{\vec{k}, \alpha} V_j(\vec{r} - \vec{R}_n - \vec{r}_\alpha) - \frac{1}{2\Omega} \sum_{\sigma = \uparrow, \downarrow} V_{\text{Coul}}(G) \rho(G) + \frac{1}{4\Omega} \sum_{\sigma = \uparrow, \downarrow} V_j(\vec{r}) \rho(G) + \frac{1}{\Omega} \sum_{\sigma = \uparrow, \downarrow} \{\epsilon_{\text{cor}}(G) - V_{\text{cor}}(G)\} \rho(G) + \sum_\alpha Z_a F_a(G = 0) + E_{\text{ii}}.
\]

where the first term is the total bare pseudopotential \( V_j(r) \) over the unit cells at \( \vec{R}_n \) and sites at \( \vec{r}_\alpha \), and \( V_{\text{Coul}} \), \( V_j(\vec{r}) \) (with an exchange coefficient of \( \alpha = \frac{3}{2} \)), and \( V_{\text{cor}} \) are the interelectronic Coulomb, exchange, and correlation\(^{17}\) screening due to the pseudodensity \( \rho(\vec{r}) \). The single-particle equation with the potential \( V_{\text{eff}}(\rho(\vec{r})) \) is solved self-consistently by expanding the wave functions of band \( j \) and wave vector \( \vec{k} \) in a mixed representation\(^{16}\) of Gaussian Bloch functions \( \Phi_{\alpha \lambda}(\xi, \vec{r}) \) (with exponents \( \xi_j \) and site index \( \alpha \)) and plane waves

\[
\psi_j(\vec{k}, \vec{r}) = \sum_{\ell \alpha} A_{\ell \alpha}(\vec{k}) \Phi_{\ell \alpha}(\xi, \vec{k}, \vec{r}) + \sum_{G \neq 0} B_j(\vec{k} + G) e^{i(\vec{k} + \vec{G}) \cdot \vec{r}}.
\]

This allows for an efficiently converged representation of both the localized pieces of the covalent charge density (via the first term) and the extended "free-electron-like" parts (via the second term). The coefficients \( A_{\ell \alpha}(\vec{k}) \) and \( B_j(\vec{k} + G) \) are determined by a linear variation while the exponents \( \xi_j \) for \( j = 0 \) and 1 are chosen by a nonlinear variation on the total energy at each lattice parameter (e.g., \( \xi_1 - \xi_0 = 1.5 \) a.u.\(^{-2} \) for \( a = 5.43 \) Å).

The total energy per cell is calculated in a momentum-space representation\(^{19}\) from the DF expression.

\[
E_{\text{tot}} = \sum_{\ell \alpha} N_j(\vec{k}) \xi_j(\vec{k}) - \frac{1}{2\Omega} \sum_{\sigma = \uparrow, \downarrow} V_{\text{Coul}}(G) \rho(G) + \frac{1}{4\Omega} \sum_{\sigma = \uparrow, \downarrow} V_j(\vec{r}) \rho(G) + \frac{1}{\Omega} \sum_{\sigma = \uparrow, \downarrow} \{\epsilon_{\text{cor}}(G) - V_{\text{cor}}(G)\} \rho(G) + \sum_\alpha Z_a F_a(G = 0) + E_{\text{ii}}.
\]

Here \( N_j(\vec{k}) \) and \( \xi_j(\vec{k}) \) are the band occupation numbers and energies, respectively; \( \epsilon_{\text{cor}} \) is the homogeneous correlation energy per particle\(^{17}\). \( F_a(G = 0) \) is the zero-momentum limit of the non-Coulombic part of the local pseudopotential (chosen here as the \( l = 0 \) component), \( E_{\text{ii}} \) is the Ewald ion-ion energy, and \( Z_a \), \( \Omega \), are, respectively, the \( \alpha \)th site valence charge (\( Z_a = 4 \) for Si) and unit-cell volume. The Fourier components of the Coulomb, exchange, and correlation potentials and the correlation energy are indicated by \( V_{\text{Coul}}(G) \), \( V_j(\vec{r}) \), \( V_{\text{cor}}(G) \), and \( \epsilon_{\text{cor}}(G) \), respectively. The self-consistent iterative procedure is carried to a consistency of \( 10^{-5} \) Ry in the potential and the Brillouin-zone integration is replaced by 10 special \( \vec{k} \) points. The plane-wave expansion in Eq. (3) is carried to an energy cutoff of 16 Ry (i.e., 283 plane waves as \( \Gamma \)). Higher-momentum components are represented by the Gaussian Bloch functions. The charge density and potential include waves up to an energy of 60 Ry. The internal precision of the total energy is \( 10^{-4} \) Ry. Within the
underlying density-functional pseudopotential approach, the only approximation made in our present method is the truncation of basis sets and reciprocal vectors sums to finite values. This, in turn, is a controlled approximation.

The components of the self-consistent potential [Eq. (2)] in Si at the equilibrium lattice constant are depicted in Fig. 2. The exchange potential is seen to effectively screen the Coulomb repulsion whereas the correlation potential is almost an order of magnitude weaker. The Coulomb and exchange screening are strongly structured and nonuniform, while the correlation screening is nearly uniform. We find that the characteristic bond-elongated form of the screening is lost when the high-momentum components of the pseudopotential are ignored [viz. in using $V_{\text{emp}}(r)$]. Instead, one obtains an ellipsoidal screening which has its long axis perpendicular to the Si–Si bond. This leads to forces that tend to push the atoms too close together.

The self-consistent pseudocharge density $\rho(\vec{r})$ is compared with the experimentally synthesized valence density$^{20}$ in Fig. 3. The details as well as the magnitude of the charge are well reproduced. Table II shows the experimental x-ray structure factors of Si (corrected for the Debye-Waller factor and anomalous dispersion$^{21}$) and the results calculated by numerically Fourier transforming the density ob-

---

**FIG. 2.** Components of the self-consistent crystal potential of Si (in Ry) in the (110) plane. $V_{\text{Coul}}(\vec{r})$, $V_{\text{ex}}(\vec{r})$, $V_{\text{cor}}(\vec{r})$, and $V_{\text{emp}}(\vec{r})$ are the interelectronic Coulomb and local pseudopotentials (both with the exclusion of the divergent $\vec{G} = 0$ term), the exchange and correlation potentials, respectively. The zero of $V_{\text{Coul}}(\vec{r})$ is set at 0.0.

**FIG. 3.** Experimental (Ref. 20) and calculated valence charge density in the (110) plane of Si (in units of electron per cell).
sidered, as our first-principles pseudopotentials eliminate the core electrons and reflect at the same time an accuracy comparable to that of the all-electron DF treatment.

An interesting issue is the relation between the calculated cohesive properties and the degree of repulsiveness of the core pseudopotential. Given the differences in the ground-state valence charge density obtained with the soft-core pseudopotential of Ref. 6 (see Ref. 12) and with the present hard-core pseudopotential, one may ask whether one could construct a better soft-core pseudopotential that will produce an accurate description of the cohesive properties.

Consider the amount of charge \( Q(R) \) of angular symmetry \( l \) which is enclosed between an atomic origin and a distance \( R \). In the present DF pseudopotential approach the core charge \( Q(R) \) (where \( R_c \) is some core radius) is determined by the condition that pseudo wave function \( \chi(r) \) be equal to the all-electron valence wave function \( \psi(r) \) from \( r = \infty \) to the smallest value \( r = R_c \) possible under the constraint that \( \chi(r) \) be nodeless, monotonic, and expressible as a linear combination of the density functional basis [Eq. (1)]. Once this "maximum similarity" criteria between \( \chi(r) \) and \( \psi(r) \) is established outside the core, the pseudo wave function \( \chi(r) \) is anchored to the atomic site at \( r = 0 \) by requiring \( \chi(0) = \chi'(0) = \chi''(0) = 0 \). This allows one to extend the spatial domain of wave function similarity (between \( r = \infty \) and \( r = R_c \)) closer to the core. At the same time this produces a hard-core pseudopotential (i.e., \( \lim_{r \to 0} V(r) = A_1/r^2 \); \( A_1 > 0 \)) with its attendant low-energy dependence (i.e., the origin dependence of the wave function is unchanged to within a good approximation by bonding forces in chemical environments which are different from the one used for constructing the pseudopotential).

If one now imagines a process in which the pseudopotential is made gradually softer in the core region [for instance, by mixing in Eq. (1) more core orbital character than the minimum amount needed to obtain a nodeless and monotonic pseudo wave function], the pseudo wave functions will increase their penetration into the core region, increasing thereby \( Q(R_c) \). As \( \int_0^{R_c} Q(R) \, dR \) is normalized, such a transfer of charge is accompanied by a decrease in the bond charge. In the soft-core pseudopotential of Ref. 6 the increase in the core charge is accomplished by peaking the wave functions at the bond center and deleting charge density between the bond center and the core region (cf. Ref. 12). As indicated above, this results in a small equilibrium lattice parameter for bulk Si. Note that shifting \( Q(R_c) \) into the core lowers the (negative) core attraction contribution \( \sum_i (x_i | V_i | x_i) \) to the total energy while peaking the charge density at the bond center increases the (positive) interelectronic Coulomb repulsion. It is hence likely that one could design a soft-
core pseudopotential with a “degree of softness” which will shift just enough charge \( Q \) into the core region to produce a desired balance between the opposing contributions to the total energy, at the observed lattice parameter. This could conceivably be achieved through the device of controlling the core orbital overmixture in the formulation of Eq. (1) (i.e., adjusting \( C_{1s}, C_{2s}, C_{2p} \), or in other procedure) by empirically adjusting the position \( R \), where the “true” orbital \( \psi_0(r) \) matches the pseudor-orbital \( \chi_0(r) \). It will then be of interest to find if this total energy is physically accurate away from the equilibrium position (e.g., the bulk modulus and elastic constants), and if the pseudopotential used is transferable to other systems. In the present DF pseudopotential approach we avoid such empirical adjustments altogether by taking the extreme limit of hard-core pseudopotentials where the pseudocharge density resembles the all-electron valence charge density over the largest spacial domain possible.

Finally, I note that although the density-functional description of exchange and correlation effects produces an accurate representation of the ground-state properties, it does not permit the identification of the associated energy eigenvalues (i.e., band structure) with the actual electronic elementary excitations in the system. Hence, for example, while both AE and the pseudopotential band structure predict a minimum energy gap of 0.6 eV for Si, the observed threshold for electronic excitation is almost twice as high. It would seem that one of the more acute problems of contemporary electronic structure theory still lies in identifying simple formulations for the effective potential which will produce physically meaningful single-particle energies without sacrificing the accuracy of the calculated ground-state properties.

---